

Selected Data Manipulation Techniques in SAS

Using SAS Studio on SAS On Demand for Academics (SODA)



Imported Your Data Already?

- If you already have your data in SAS Studio on SAS On Demand for Academics, you can skip the slides providing an overview of the import process.



Dataset

- This tutorial is a walkthrough with a sample set of data. You may use this to walk through the tutorial, if you wish, but for your assignments, you will be asked to use your own dataset (as specified within the course).

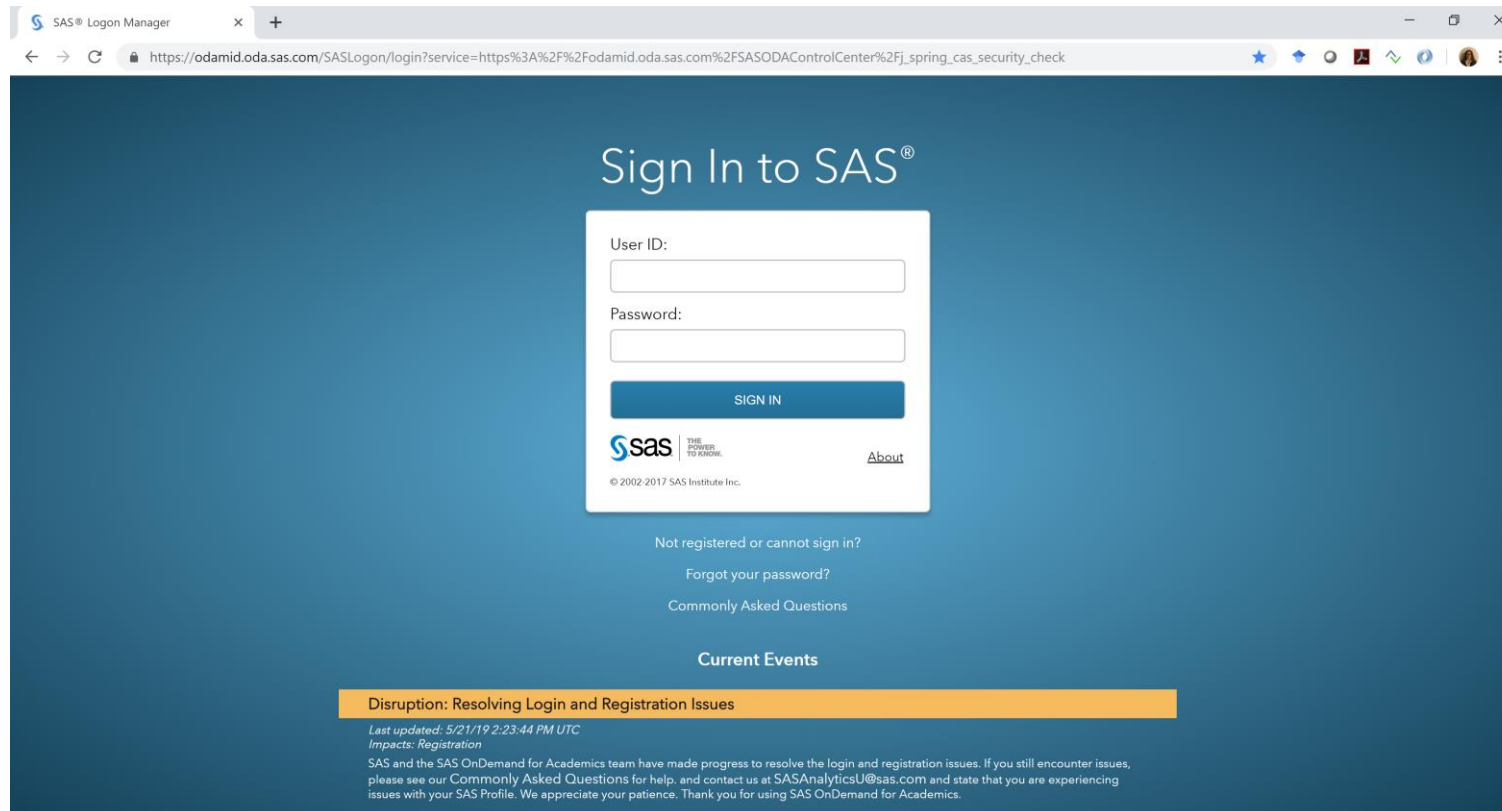
Dataset reference:

Skoryk, M. (2021). Sepsis Prediction from Clinical Data. Version 1.
Retrieved from <https://www.kaggle.com/maxskoryk/datasepsis>



Access the SAS OnDemand for Academics Control Center

<https://odamid.oda.sas.com/SASODAControlCenter>



SAS OnDemand for Academics (SODA) Control Center

The screenshot shows the SAS OnDemand for Academics (SODA) Control Center dashboard. The browser address bar displays `https://odamid.oda.sas.com/SASODAControlCenter/`. The page header includes the SAS logo and the text "SAS OnDemand for Academics Control Center - United States". A blue navigation bar at the top contains the word "Dashboard" and a user profile for "Stefanie Reay".

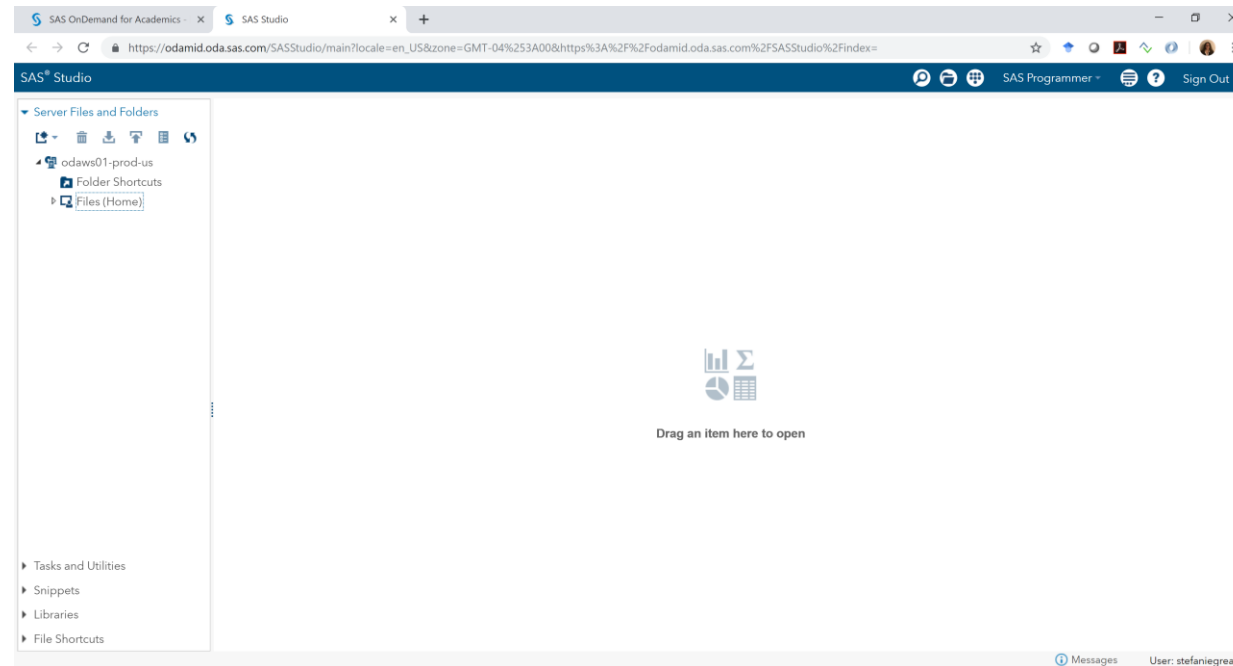
The main content area is divided into several sections:

- Current Events** (click entry for details):
 - Resolving Login and Registration Issues**: Last updated: May 21, 2019 14:23 UTC. Impacts: Registration, Login Page.
- Planned Events** (click entry for details):
 - Scheduled Maintenance on Tuesday, July 16, 8:30 - 10:00 am EDT**: Last updated: Jul 09, 2019 17:38 UTC. Scheduled start time: Jul 16, 2019 12:30 UTC. Scheduled end time: Jul 16, 2019 14:00 UTC.
- Applications** (tabbed view):
 - SAS Studio**: Write and run SAS code with a Web-based SAS development environment. Actions: Clear my saved tabs.
 - SAS Enterprise Guide**: Deliver the power of SAS from an easy-to-use, point-and-click interface. (Download Required)
 - JMP access to SAS Servers (U.S. users only)**: Statistical discovery software. Users must have a copy of JMP software. (Configuration Steps Required)
 - SAS Enterprise Miner**: Reveal valuable insights with powerful data mining software. (Configuration Steps Required)
- Reference**:
 - Support Site
 - Step-by-Step Registration Guides
 - User's Guide
 - Commonly Asked Questions
 - Status Page
- Quotas** (learn more):
 - Home Directory (4.4MB/5120MB): 0%
 - Course Directory (207.0MB/3072MB): 7%
- Notices**:
 - German, Spanish and French Locale Users Impacted by a SAS Studio Issue**: Last updated: Jun 18, 2019 15:02 UTC. Importance: High. Impacts: SAS Studio.

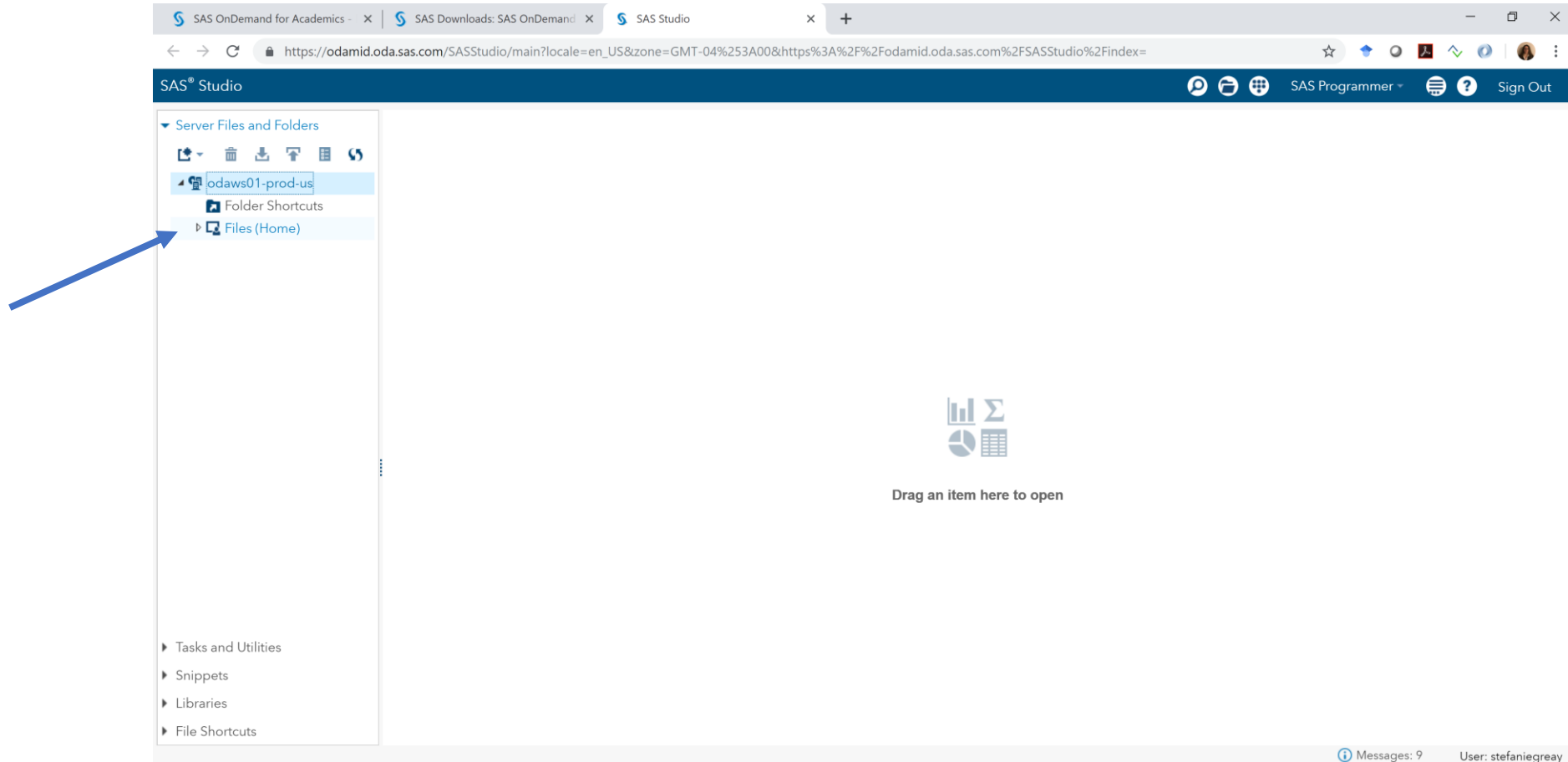


SAS Studio

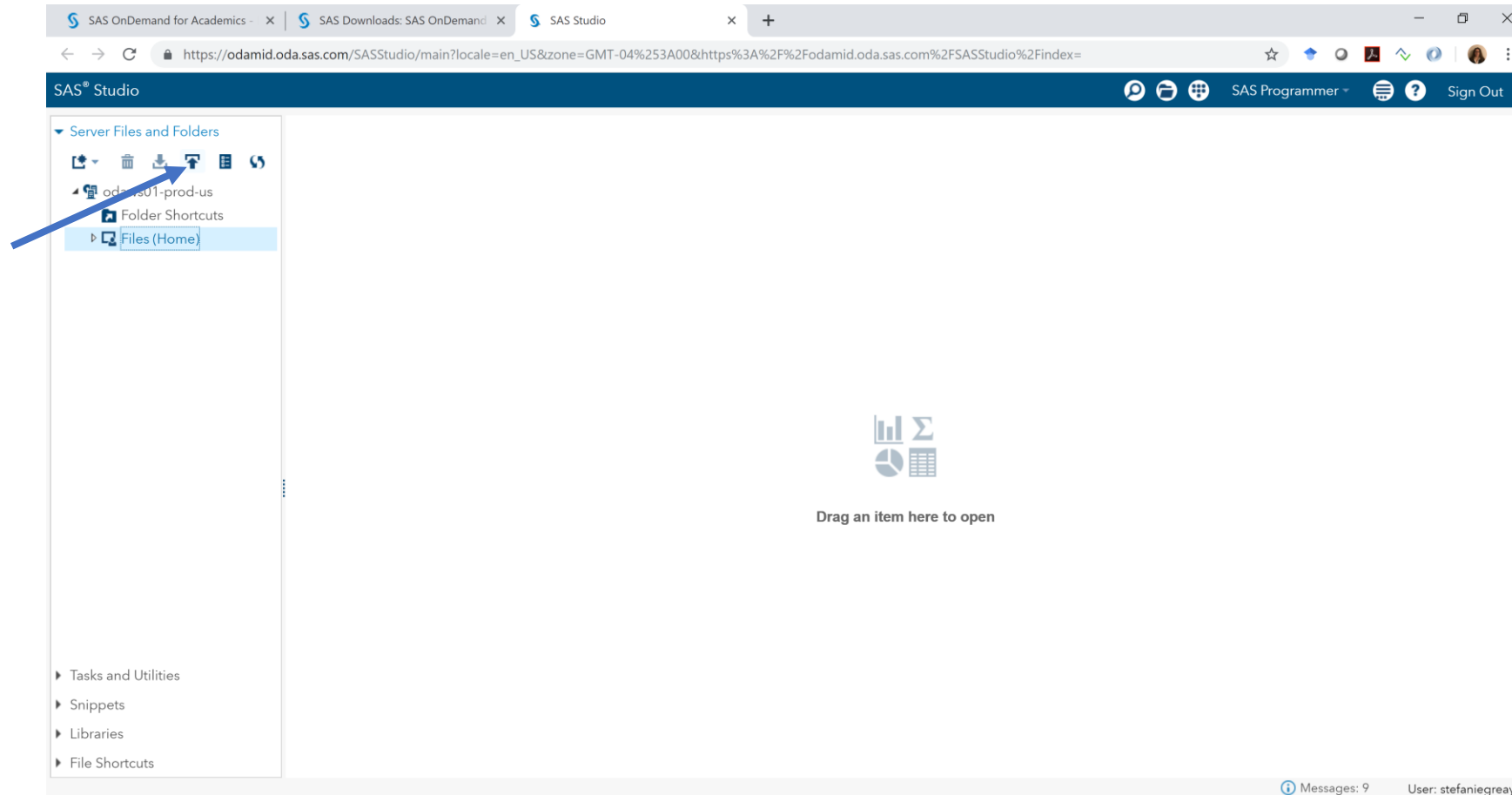
<https://odamid.oda.sas.com/SASStudio/>



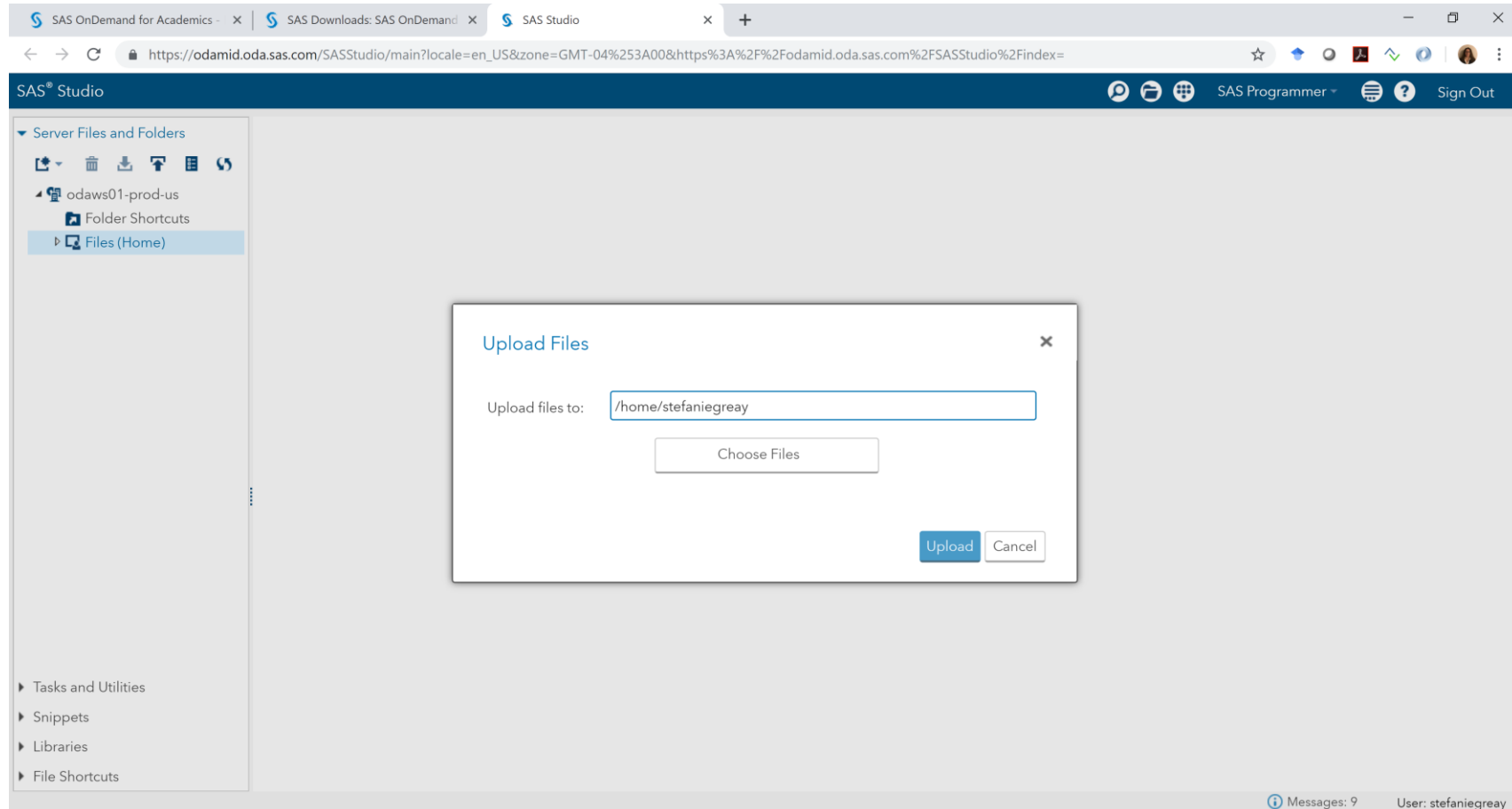
Click on Files(Home)



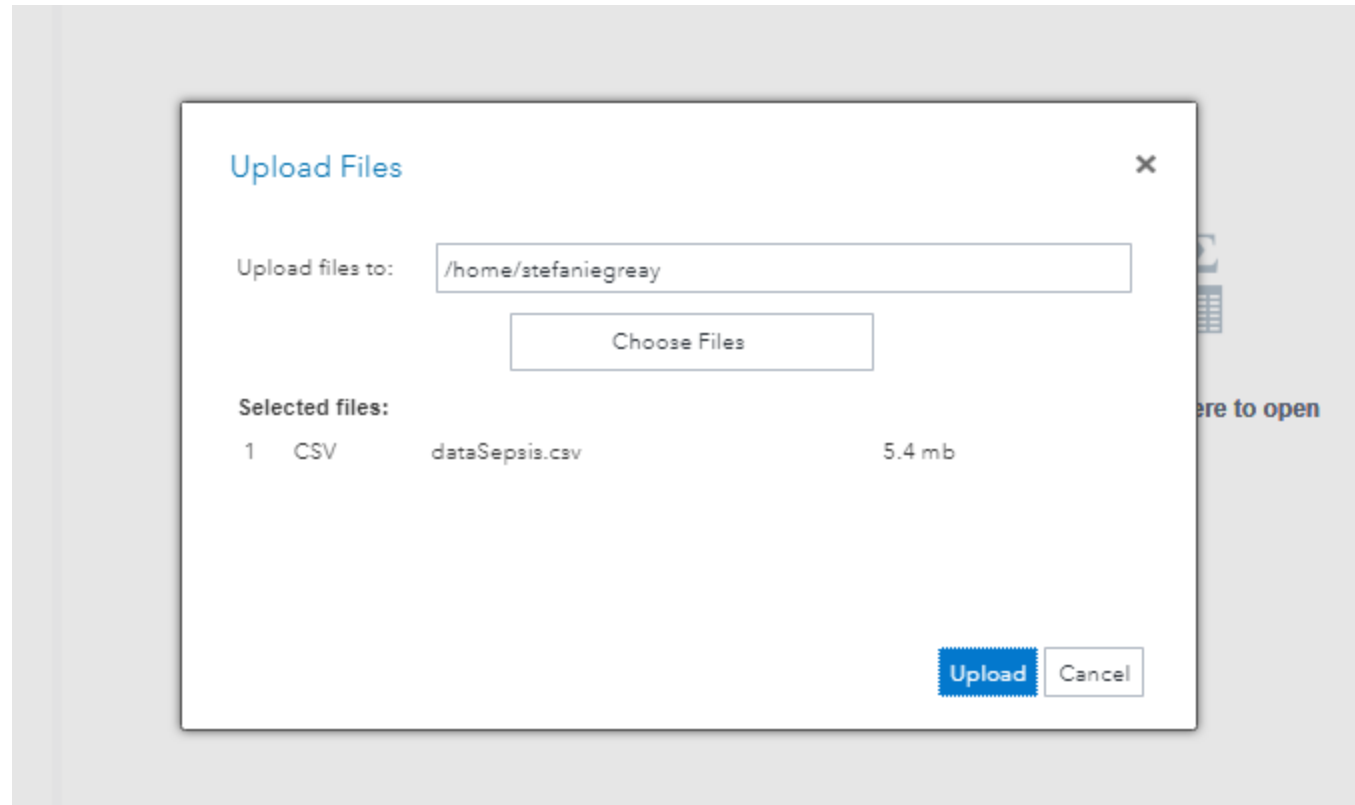
The Upload button will display in dark blue



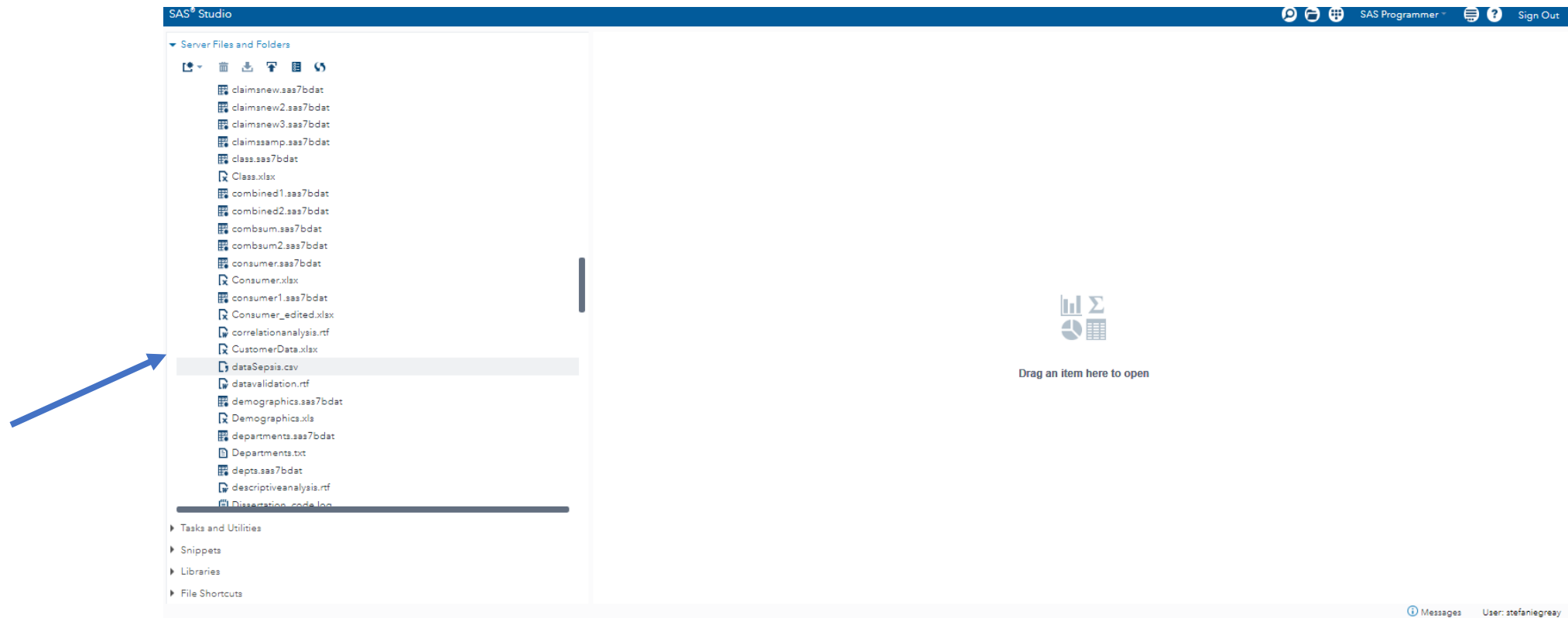
You can create a folder at this point, if you wish, or simply upload to your home directory.



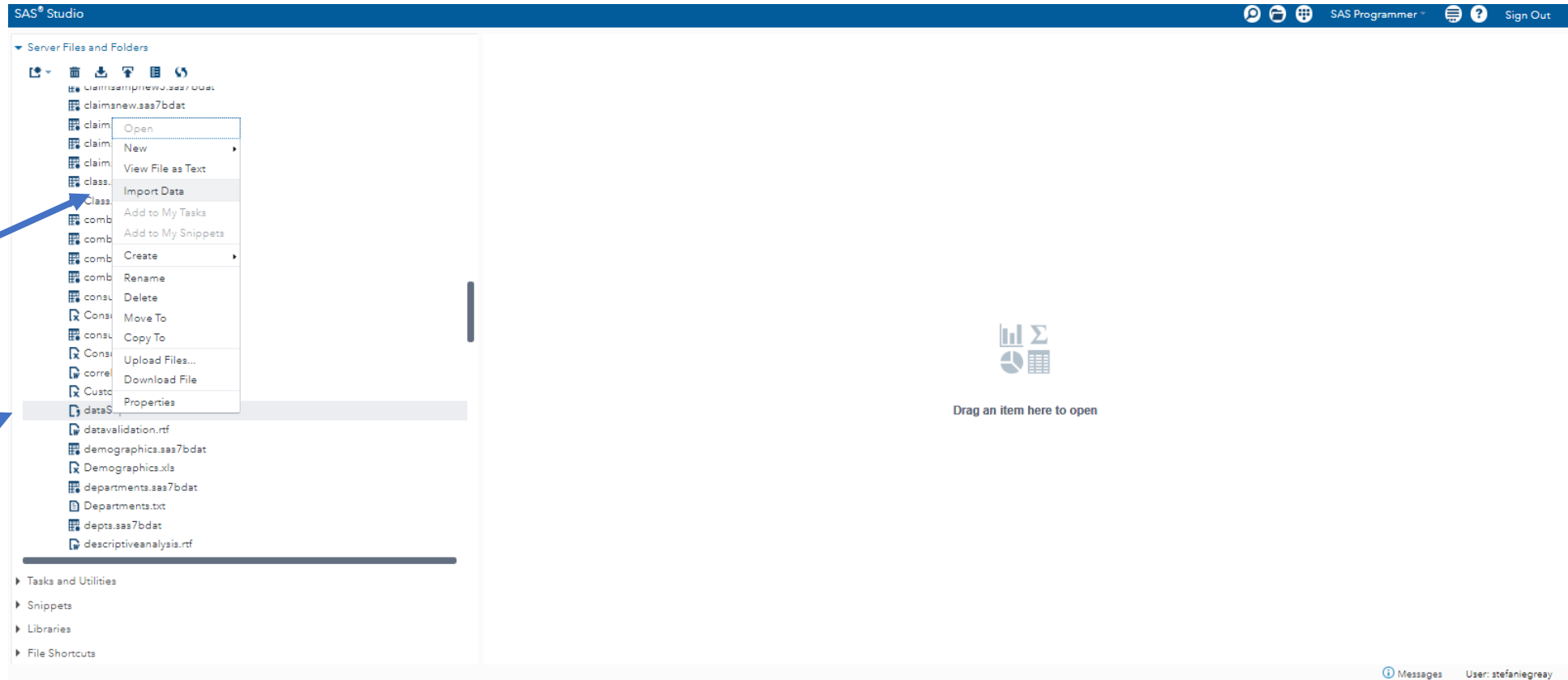
Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”



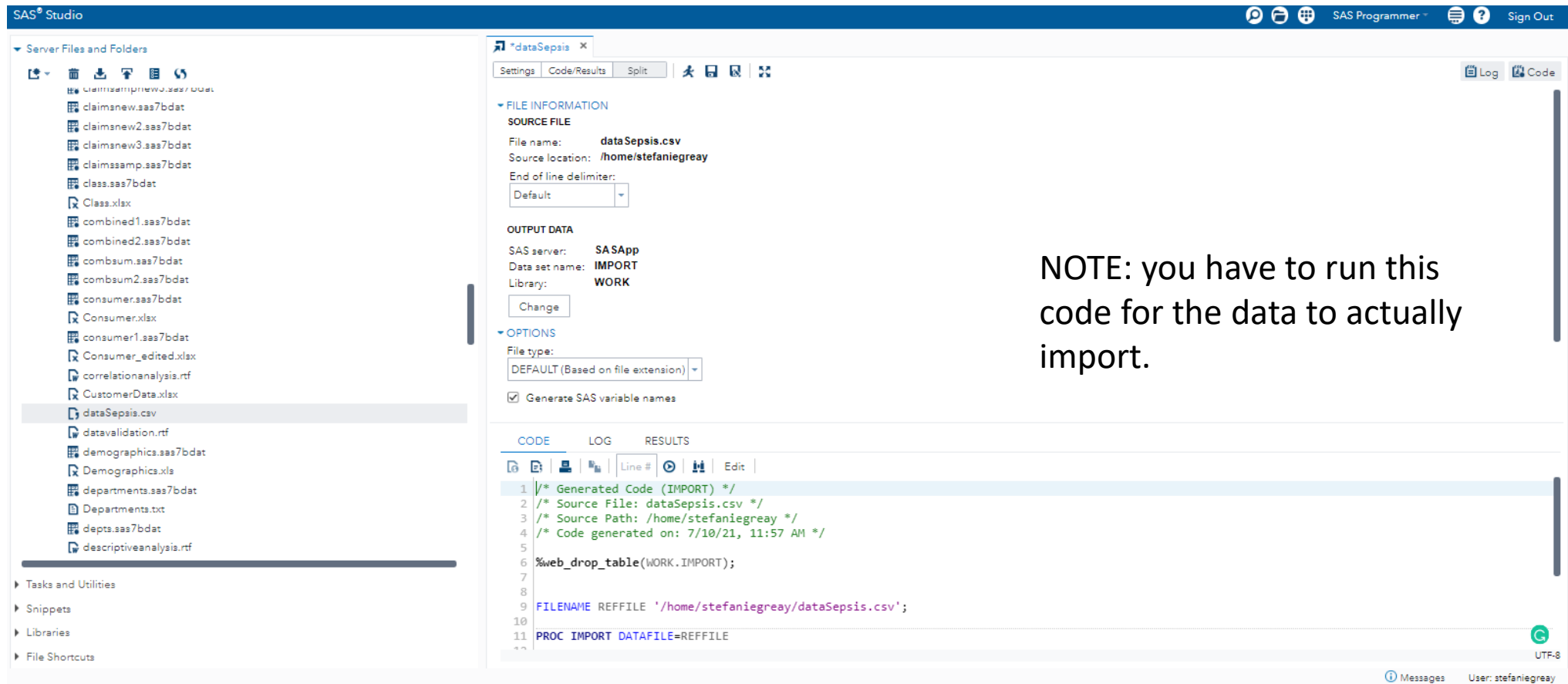
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



To import the dataset into a SAS dataset format (from the current csv format), right click on the name of the file, and select “Import Data.”



The Proc Import code will be written for you (save this as a template to use for future imports!)



NOTE: you have to run this code for the data to actually import.

```
1 /* Generated Code (IMPORT) */
2 /* Source File: dataSepsis.csv */
3 /* Source Path: /home/stefaniegreay */
4 /* Code generated on: 7/10/21, 11:57 AM */
5
6 %web_drop_table(WORK.IMPORT);
7
8
9 FILENAME REFFILE '/home/stefaniegreay/dataSepsis.csv';
10
11 PROC IMPORT DATAFILE=REFFILE
```



The Proc Import code will be written for you (save this as a template to use for future imports!)

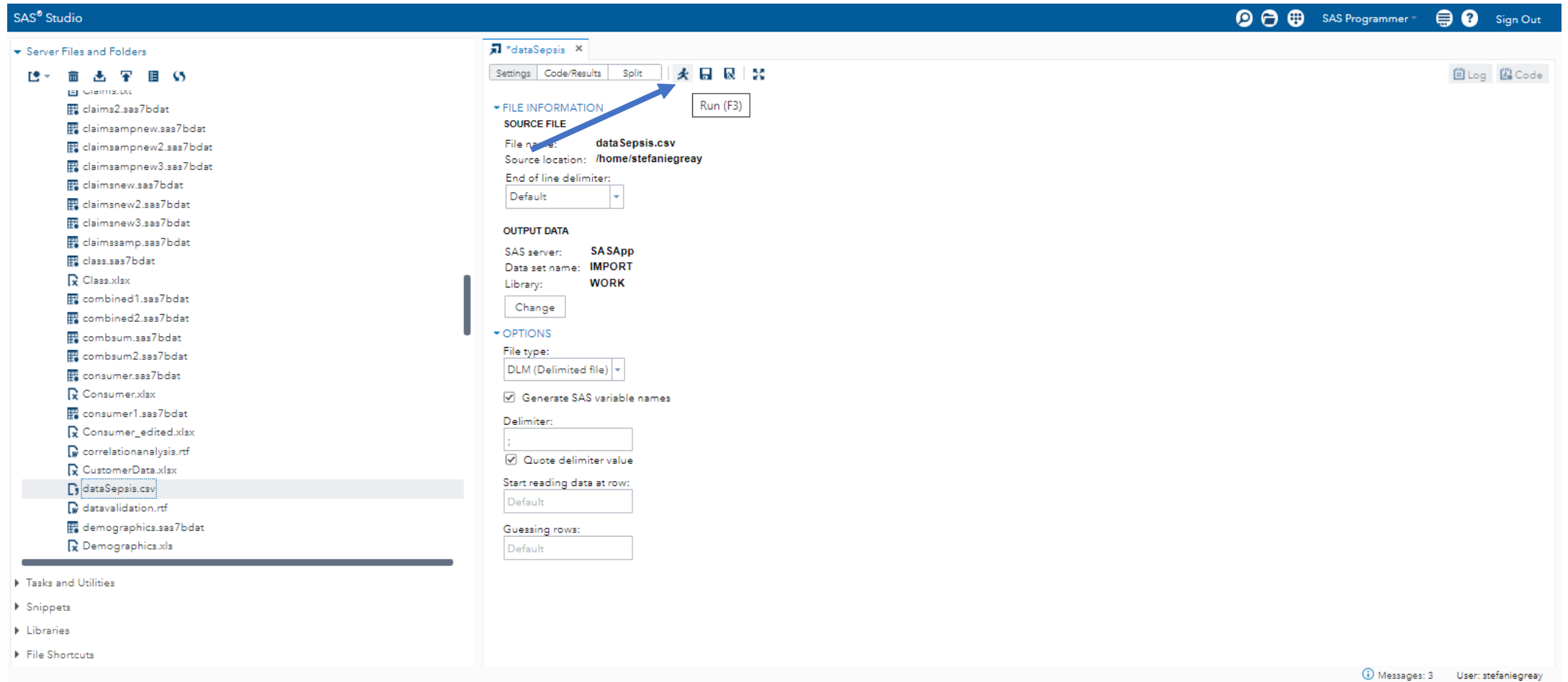
The screenshot shows the SAS Studio interface. On the left is the 'Server Files and Folders' pane with a list of files including 'dataSepsis.csv'. The main window displays the 'Proc Import' wizard for 'dataSepsis.csv'. The 'FILE INFORMATION' section shows the source file name and location. The 'OUTPUT DATA' section shows the SAS server, data set name 'IMPORT', and library 'WORK'. The 'OPTIONS' section shows 'File type' as 'DLM (Delimited file)', 'Generate SAS variable names' checked, 'Delimiter' as ';', 'Quote delimiter value' checked, 'Start reading data at row' as 'Default', and 'Guessing rows' as 'Default'.

NOTE: you have to run this for the data to actually import.

These options shown here are the appropriate ones for the sepsis sample data. They will likely have to be adjusted for your dataset, although the default settings might work just fine for yours.



To run the code, click the icon that looks like a guy running.



When you run the import, you will see the dataset and summary in the output data window when you click the “Code/Results” or “Split” tab and then “Output Data” and can verify its success.

SAS® Studio

Server Files and Folders

- class.sas7bdat
- Class.xlsx
- combined1.sas7bdat
- combined2.sas7bdat
- combsum.sas7bdat
- combsum2.sas7bdat
- consumer.sas7bdat
- Consumer.xlsx
- consumer1.sas7bdat
- Consumer_edited.xlsx
- correlationanalysis.rtf
- CustomerData.xlsx
- dataSepsis.csv**
- datavalidation.rtf
- demographics.sas7bdat
- Demographics.xls
- departments.sas7bdat
- Departments.txt
- depts.sas7bdat
- descriptiveanalysis.rtf
- Dissertation_code.log
- Dissertation_code.sas
- Dissertation_code_01062020_243pm.log
- Dissertation_code_01062020_243pm.sas
- Dissertation_code_01082020_1121am.sas

Tasks and Utilities

- Snippets
- Libraries
- File Shortcuts

*dataSepsis

Settings | Code/Results | Split

CODE LOG RESULTS **OUTPUT DATA**

Table: WORK.IMPORT View: Column names Filter: (none)

Columns

- Select all
- ☒ HR
- ☒ O2Sat
- ☒ Temp
- ☒ SBP
- ☒ MAP
- ☒ DBP
- ☒ Resp
- ☒ EtCO2
- ☒ BaseExcess
- ☒ HCO3
- ☒ FiO2
- ☒ pH

Property Value

Total rows: 36302 Total columns: 41

	HR	O2Sat	Temp	SBP	MAP	DBP	Resp	EtCO2	BaseExcess	HCO3	FiO2	pH	PaCO2	SaO2	AST	BUN	Alkalinephos	Cr
1	103	90	NaN	NaN	NaN	NaN	30	NaN	21	45	NaN	7.37	90	91	16	14	98	9.
2	58	95	36.11	143	77	47	11	NaN	NaN	22	NaN	NaN	NaN	NaN	NaN	100	NaN	7.
3	91	94	38.5	133	74	48	34	NaN	NaN	31	0.8	NaN	NaN	NaN	NaN	30	NaN	10.
4	92	100	NaN	NaN	NaN	NaN	NaN	NaN	NaN	29	NaN	NaN	NaN	NaN	NaN	9	NaN	Na
5	155.5	94.5	NaN	147.5	102	NaN	33	NaN	-12	13	1	7.22	36	NaN	452	68	88	5.
6	73	99	36.06	100	67	49.5	16.5	NaN	-8	16	NaN	7.27	37	NaN	NaN	28	NaN	7.
7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	25	NaN	7.35	48	NaN	NaN	NaN	NaN	Ni
8	82	100	35.5	112	79.5	63	14	NaN	0	23	1	7.42	37	NaN	NaN	18	NaN	Ni
9	89	100	NaN	141	85	57	17	NaN	1	25	NaN	7.43	37	NaN	NaN	9	NaN	8.
10	100	95	37.28	121	20	NaN	NaN	NaN	NaN	22	NaN	NaN	NaN	NaN	NaN	32	NaN	7.
11	95	100	NaN	89	62.33	NaN	18	NaN	NaN	22	NaN	NaN	NaN	NaN	8	19	70	7.
12	86	96	38	111	66	49	17	NaN	1	27	NaN	7.39	45	95	NaN	16	NaN	Ni
13	88	100	36.3	99	66	52	16	NaN	-3	20	1	7.35	39	NaN	NaN	14	NaN	Ni
14	116	97	38.28	200	108	90	24	NaN	6	NaN	0.7	7.51	39	NaN	NaN	NaN	NaN	Ni
15	110	99	36.4	116	219	66	19	NaN	-8	19	NaN	7.22	46	96	NaN	85	NaN	Ni
16	54	95	NaN	103	63	NaN	11	NaN	NaN	30	NaN	NaN	NaN	NaN	NaN	11	NaN	9.
17	98	94	NaN	95	62	45	15	NaN	NaN	26	NaN	NaN	NaN	NaN	12	11	55	7.
18	72	96	NaN	103	62	45	20	NaN	-1	NaN	NaN	7.4	36	98	NaN	NaN	NaN	Ni
19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	24	NaN	NaN	NaN	NaN	NaN	65	NaN	9.
20	84	98	NaN	106	67.33	NaN	29	NaN	NaN	20	NaN	NaN	NaN	NaN	NaN	37	NaN	6.
21	69.5	100	37.17	109.5	71	NaN	17.5	NaN	NaN	24	NaN	NaN	NaN	NaN	NaN	14	NaN	Ni

Messages: 4 User: stefaniegreay



When you click the “Code/Results” or “Split” tab and then “Results,” you can see the contents of the dataset, to verify the number of observations and variables are as expected.

The screenshot shows the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'dataSepsis.csv' selected. The main window displays the 'Results' tab for the 'dataSepsis' dataset. The 'Table of Contents' section shows 'The CONTENTS Procedure' output, which includes a table of dataset properties and a table of engine/host dependent information.

The CONTENTS Procedure			
Data Set Name	WORK.IMPORT	Observations	36302
Member Type	DATA	Variables	41
Engine	V9	Indexes	0
Created	07/10/2021 13:19:25	Observation Length	168
Last Modified	07/10/2021 13:19:25	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

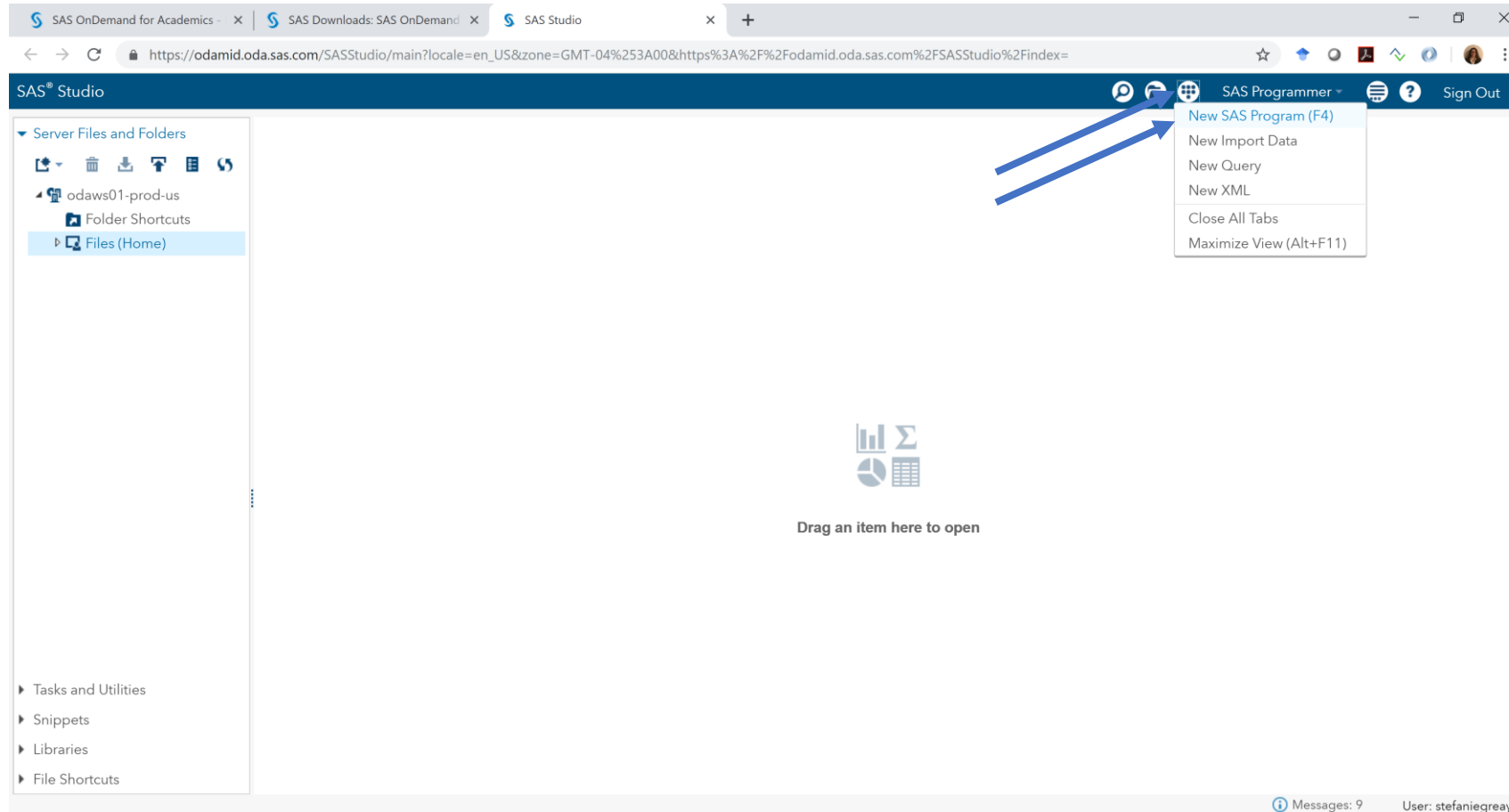
Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	47
First Data Page	1
Max Obs per Page	779
Obs in First Data Page	736
Number of Data Set Repairs	0
Filename	/sas/work/SAS_work6E5F0000DA42_odswe03-usw2.ods.sas.com/SAS_work75820000DA42_odswe03-usw2.ods.sas.com/Import.sas7bdat
Release Created	9/04/2016
Host Created	Linux
Inode Number	537069780
Access Permission	rw-r--r--
Owner Name	stefaniegreay
File Size	6MB
File Size (bytes)	6291456

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
15	AST	Char	3	\$3.	\$3.
35	Age	Num	8	BEST12.	BEST32.
17	Alkalinephos	Char	3	\$3.	\$3.

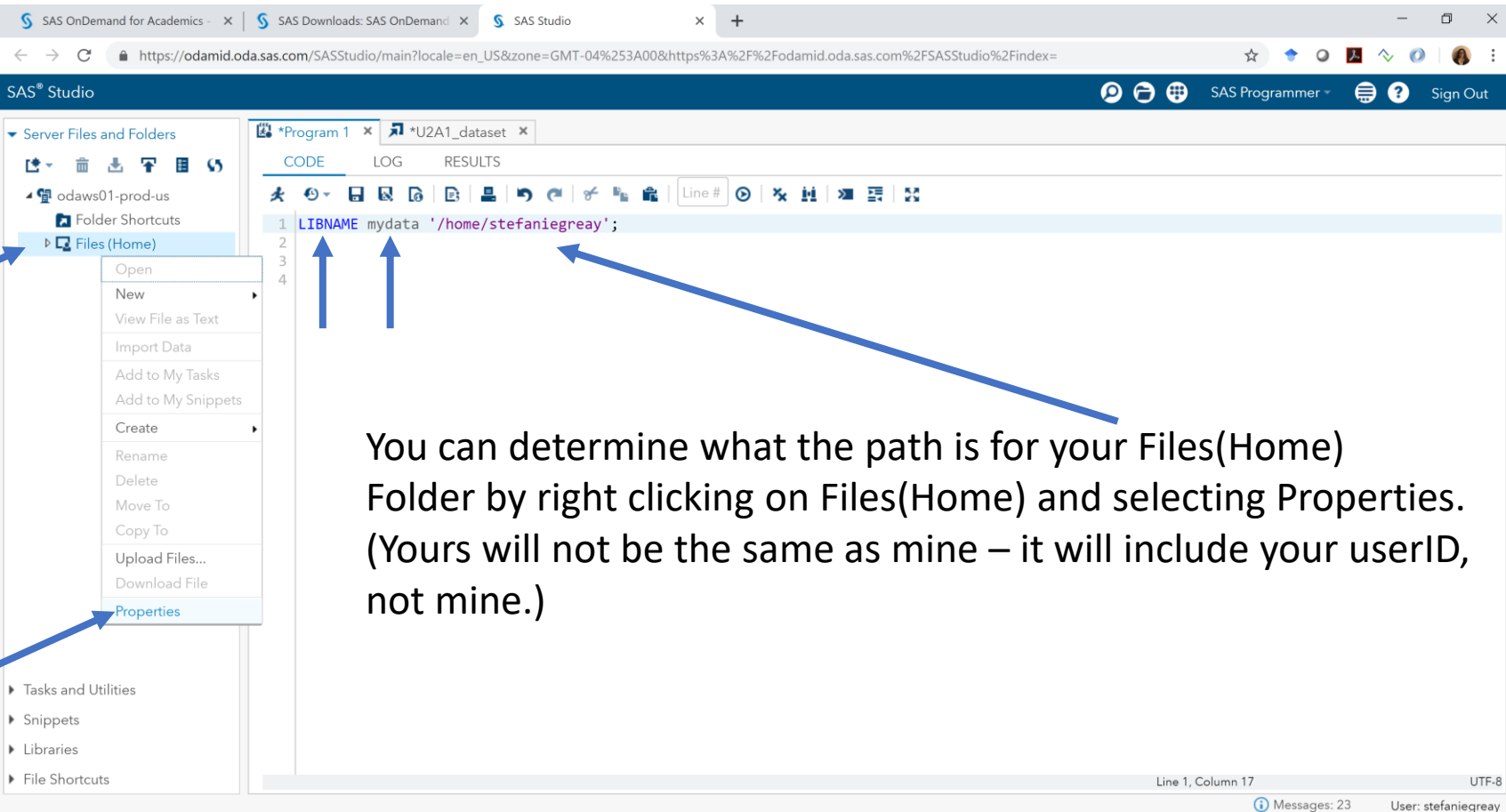
At the bottom right of the interface, it shows 'Messages: 4' and 'User: stefaniegreay'.



To get started working with the dataset you just imported, start a new SAS program.



To create a SAS Library for your Files(Home) folder, you need to use a libname statement



The screenshot shows the SAS Studio web interface. On the left, the 'Server Files and Folders' pane displays a tree structure with 'odaws01-prod-us' expanded, showing 'Folder Shortcuts' and 'Files (Home)'. A right-click context menu is open for 'Files (Home)', with the 'Properties' option highlighted at the bottom. A blue arrow points from the 'Files (Home)' folder to the 'Properties' option. In the center, the 'CODE' pane shows a SAS program with the following code on line 1:

```
1 LIBNAME mydata '/home/stefaniegreay';
```

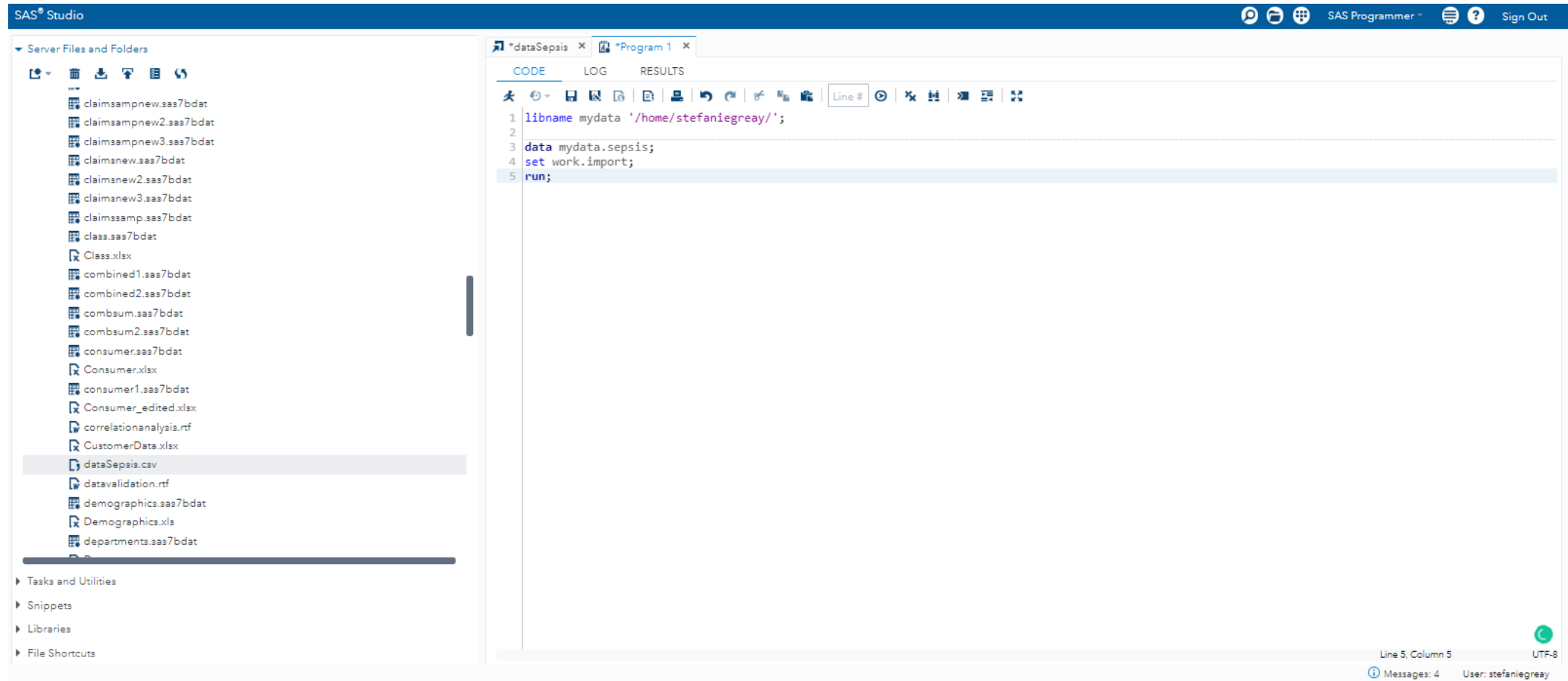
Two blue arrows point to the components of this statement: one to 'LIBNAME' and another to 'mydata'. A third blue arrow points from the 'Properties' option in the context menu to the path '/home/stefaniegreay' in the code. Below the code pane, a text box explains the path determination process.

You can determine what the path is for your Files(Home) Folder by right clicking on Files(Home) and selecting Properties. (Yours will not be the same as mine – it will include your userID, not mine.)

The bottom status bar of the SAS Studio window shows 'Line 1, Column 17', 'UTF-8', 'Messages: 23', and 'User: stefaniegreay'.



Save the temporary SAS dataset created by the import to your library using the following sample code.



Data Step

```
1 data mydata.sepsis2;
2 set mydata.sepsis;
3 Gender_c=input(Gender,$6.);
4 if Gender=1 then Gender_c='Male'
5 else Gender_c='Female';
6 drop Gender;
7 run;
```

Data statement:
Specifies the output library and dataset

Set statement:
Specifies the input library and dataset

Data manipulation statements:
These statements are where the data can
be formatted, reformatted, calculated,
replaced, manipulated, etc.

Run statement:
All procedures and data steps in SAS end
with a run statement.



Selected Data Manipulation Examples

- Formatting a variable
- Converting a numeric variable to a character variable
- Converting a character variable to a numeric variable
- Creating a calculated variable
- Populating values of a variable using an if-then statement
- Combining multiple character variables
- Taking only a portion of a character variable using a substring



Formatting a variable

```
4 | format Unit2 2.;
```



format function:
Tells SAS that it to format a
specified variable using the
specified format

Variable to
be
formatted

Format:
Tells SAS how to format the variable.
2. is a numeric format that is exactly 2 places
long (specified by the 2) and the period ends
the format (the period is on all informats and
formats)



Converting a numeric variable to a character variable

```
4 Gender_c=input(Gender,$6.);
```

New variable name

Input function:
Tells SAS that it will be importing the values from one variable to another variable, using the informat specified

Old variable name

Informat:

Tells SAS what format to import the values of the variable in as.
\$6. is a character format (specified with the \$) that is 6 characters long (specified by the 6) and the period ends the format (the period is on all informats and formats)



Converting a character variable to a numeric variable

```
+  
;  
Unit2_n=input(Unit2,best12.);  
-
```

New variable
name

Input function:
Tells SAS that it will be importing
the values from one variable to
another variable, using the
informat specified

Old variable
name

Informat:
Tells SAS what format to import the values of
the variable in as.
best12. is a numeric format that is up to 12
places long (specified by the 12) but uses the
best format (specified by the best) and the
period ends the format (the period is on all
informat and formats)



Creating a calculated variable

```
20  
21 Unitsum=Unit1_n + Unit2_n;  
22
```

New variable
name

Variable 1 to be added

Mathematical operator
(This can be any usual
mathematical operator (+ - /
* ^ for a few examples)

Variable 2 to be added

*Calculations can contain multiple variables and multiple operators and functions enclosed in parentheses. See SAS's technical documentation for additional function options that can be used in calculations.



Populating values of a variable using an if-then statement

```
4  
5 if Calcium='NaN' then Calcium=' '  
6
```

Conditional if-then statement

Condition:

What has to be true for the following statement to execute

Result:

What will be executed if the previous condition is met

*Note that in the sepsis dataset, null values are represented by NaN by default, but in SAS, null values in character variables are represented by a blank (like that enclosed in single quotes above) and null values in numeric variables are represented by a . (i.e. a period). So this if-then statement replaces all NaN values with the correct SAS null value of a blank.



Combining Multiple Character Variables

```
66 | UnitpairID = catx(' ', Unit1, Unit2);
```

New variable
name

Concatination function

Separator (the symbol or
character that will be used
to separate the values
within the new variable)

Variables to be
combined

* Concatinated variables can contain two or more other variables. Note that a length statement might need to preceed the catx statement in order to ensure that the length of the new variable is long enough to include all of the concatenated values/variables.



Taking Only a Portion of a Character String

```
68 | subUnitPairID=substr(UnitpairID,1,2);
```

New variable
name

Substring function

Variable we are taking the
portion of the characters
from

Which
position
character
to start
with

How many
characters
to take

